Ecological Applications, 10(5), 2000, pp. 1497–1516

FIRE FREQUENCY IN THE INTERIOR COLUMBIA RIVER BASIN: BUILDING REGIONAL MODELS FROM FIRE HISTORY DATA

DONALD MCKENZIE,[1] DAVID L. PETERSON,[2] AND JAMES K. AGEE[1]

[1]*College of Forest Resources, Box 352100, University of Washington, Seattle, Washington 98195-2100 USA*
[2]*USGS Forest and Rangeland Ecosystem Science Center, Cascadia Field Station, Box 352100, Seattle, Washington 98195-2100 USA*

Abstract. Fire frequency affects vegetation composition and successional pathways; thus it is essential to understand fire regimes in order to manage natural resources at broad spatial scales. Fire history data are lacking for many regions for which fire management decisions are being made, so models are needed to estimate past fire frequency where local data are not yet available. We developed multiple regression models and tree-based (classification and regression tree, or CART) models to predict fire return intervals across the interior Columbia River basin at 1-km resolution, using georeferenced fire history, potential vegetation, cover type, and precipitation databases. The models combined semiqualitative methods and rigorous statistics. The fire history data are of uneven quality; some estimates are based on only one tree, and many are not cross-dated. Therefore, we weighted the models based on data quality and performed a sensitivity analysis of the effects on the models of estimation errors that are due to lack of cross-dating. The regression models predict fire return intervals from 1 to 375 yr for forested areas, whereas the tree-based models predict a range of 8 to 150 yr. Both types of models predict latitudinal and elevational gradients of increasing fire return intervals. Examination of regional-scale output suggests that, although the tree-based models explain more of the variation in the original data, the regression models are less likely to produce extrapolation errors. Thus, the models serve complementary purposes in elucidating the relationships among fire frequency, the predictor variables, and spatial scale. The models can provide local managers with quantitative information and provide data to initialize coarse-scale fire-effects models, although predictions for individual sites should be treated with caution because of the varying quality and uneven spatial coverage of the fire history database. The models also demonstrate the integration of qualitative and quantitative methods when requisite data for fully quantitative models are unavailable. They can be tested by comparing new, independent fire history reconstructions against their predictions and can be continually updated, as better fire history data become available.

Key words: *coarse scale; Columbia River Basin; cover types; fire effects; fire frequency; fire history; fire return interval; potential vegetation; semiqualitative methods; tree-based model.*

INTRODUCTION

Predicting the occurrence and effects of broad-scale disturbances, particularly fires, will be an important challenge for scientists and resource managers in coming decades. Significant changes in fire severity and fire size are predicted for many ecosystems as a result of land-use change, climatic change, and fire exclusion (Green 1989, Turner et al. 1989, Agee 1994, Habeck 1994, Baker 1995). Although large-scale vegetation change is constrained primarily by climate (Woodward 1987, Woodward and McKee 1991), change in fire regimes in response to climatic change could significantly alter vegetation patterns, because fire often provides critical constraints on vegetation (Fosberg et al. 1992, Baker 1995, Neilson 1995, McKenzie et al. 1996*a*). Thus, simulation models used to predict broad-scale vegetation change need to incorporate fire effects.

Manuscript received 17 December 1998; revised 20 September 1999; accepted 1 October 1999.

Most empirical research on the ecological effects of fire has been conducted at the stand level, but conclusions are often extrapolated to broader scales (McKenzie et al. 1996*b*).

Mechanistic, or process-based models are typically used for simulating fire effects (Schmoldt et al. 1999). Most of these have been constructed to represent stand-level processes and assume homogeneity of crucial inputs over the spatial scale to which they are applied (Rothermel 1972, van Wagner 1977, 1993, Kercher and Axelrod 1984, Peterson and Ryan 1986, Keane et al. 1989, 1994). Spatially explicit mechanistic models that are applied at broader scales require large amounts of empirical data as inputs (e.g., Keane et al. 1996*c*, Finney 1998), and are sensitive to the scale of resolution to which the raw data are aggregated. Also, error propagation rapidly becomes problematic in complex natural systems (Cale 1995, Pahl-Wostl 1995). In particular, spatial heterogeneity is a significant source of aggregation error through its effects on disturbance spread, fire severity, and aspects of landscape pattern (Kessell 1976, Baker 1989, Green 1989, Turner and Romme 1994).

The resolution, extent, and spatial pattern of empirical data often limit the ability to aggregate spatial data for modeling fire effects (McKenzie 1998). Geostatistical interpolation (Isaaks and Srivastava 1989) and other methods of mapping in the presence of spatial autocorrelation such as structure functions (Legendre 1993) or stochastic simulation (Rossi et al. 1993) require covariance properties that are often lacking in coarse-scale empirical data. However, empirical data generally are not adequate for process-based modeling across broad spatial scales (McKenzie et al. 1996*b*, Schmoldt et al. 1999). Coarse-scale modeling may require semiqualitative methods, including qualitative assessments of aggregation errors (Schmoldt and Rauscher 1995, Keane et al. 1996*a*, McKenzie 1998).

Fire frequency is a basic parameter in simulation models of fire effects on vegetation. It may be fixed at the beginning of model runs (Kercher and Axelrod 1984, Keane et al. 1989), or sampled at random from a probability distribution (Baker 1995, Boychuk et al. 1997). Fire frequency reconstructions provide local information, and models that use them as baseline data assume homogeneity of fire frequency over long temporal scales and over the geographic range at which they are applied.

Assembling fire history data, like most field-based research, is expensive and time consuming. There are several methods for establishing mean or median fire return intervals, and each has a different expected value for the same raw data (Agee 1993, Johnson and Gutsell 1994). The method of choice usually is determined by specific local objectives. Thus, there is a lack of consistency of methods and quality among fire history studies, and the grain and extent of studies vary significantly (Heyerdahl et al. 1995).

The dominant vegetation in forest ecosystems is often very sensitive to changes in the mean, variance, and distribution of fire return intervals. For example, forest development proceeds along different successional pathways in response to different sequences of time-since-fire (Cattelino et al. 1979, Frelich and Reich 1995, Clark 1996). Thus local information about fire frequency distributions is critical to maintain accuracy in dynamic fire modeling. Fire regimes in western North America appear to vary along gradients of temperature and moisture stress (Agee 1993). Gradients vary with the steepness of topography and degree of terrain dissection, producing significant variability in fire frequency and severity within vegetation types (e.g., Agee et al. 1990, Morrison and Swanson 1990). At meso- and fine scales, it is difficult to associate fire regimes closely with particular vegetation types in forest classifications. However, broad-scale differences in fire frequency are evidently associated with different geographic areas and distinct environmental conditions (Agee 1993, Hann et al. 1997, 1998). A coarse-scale model of fire frequency should incorporate sources of variability at multiple scales, including large-scale trends, if any, associated with latitude and longitude, climatic and elevational gradients, and any variation that can be associated with different vegetation types. The range of spatial and temporal scales will depend on the resolution and extent of variables in the model database.

In this paper, we present statistical models for predicting coarse-scale patterns of fire frequency in the Interior Columbia River basin (ICRB), using a fire history database (hereafter FHDB) from the western United States (Heyerdahl et al. 1995). The ICRB coarse-scale assessment has produced regional maps of fire regimes (both frequency and severity) for the ICRB (Keane et al. 1996*a*, Morgan et al. 1996, Hann et al. 1997). Predicted fire frequency in these maps is closely linked to the vegetation classifications produced by the ICRB assessment; specifically, broad ranges of fire frequency are predicted for each type in the classifications (Morgan et al. 1996). We take a different approach, using empirical methods to test the relationships between fire frequency and both vegetation types and environmental gradients. At most stages in the modeling process, the quality and quantity of data permitted the use of rigorous statistical methods; at one point, a heuristic approach was necessary to reconcile qualitative vegetation classifications with numeric variables.

Our principal objectives are to evaluate the effectiveness of different modeling strategies for extrapolating model results to broad spatial scales, and to examine the sensitivity of model predictions and interpretation to uncertainties in the databases. In the process, we develop fire frequency coverages for forested areas of the ICRB. Although fire severity is an equally important component of fire regimes, and has been the subject of coarse-scale modeling efforts (Lenihan et al. 1998), quantitative modeling of fire severity requires a totally different approach, and we do not address it here. We discuss the applicability of our methods to the problem of modeling coarse-scale fire effects, the advantages and disadvantages of our models compared to those from the ICRB assessment, and potential improvements in models and databases that would make broad-scale predictions more accurate.

Methods

Study area

Quigley et al. (1996) have defined the ICRB as those portions of the Columbia River Basin inside the United States east of the crest of the Cascade Mountains in Washington and Oregon, and portions of the Klamath River Basin in California and the Great Basin in Oregon, Utah, and Nevada. The ICRB covers more than 58 million ha, 46% of which is in forested vegetation. Dominant tree species range from ponderosa pine (*Pi-*

nus ponderosa) and limber pine (*Pinus flexilis*) at lower treeline to mountain hemlock (*Tsuga mertensiana*), Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), and whitebark pine (*Pinus albicaulis*) at upper treeline. Elevation of the forested areas ranges from 50 to 3700 m, and mean annual precipitation ranges from 130 to 3500 mm. Climate in the ICRB is a result of the interaction of three air masses: 1) moist marine air from the west, 2) continental air from the east and south, and 3) dry arctic air from the north (Ferguson 1997). Summer drought, caused by a seasonal northward shift in the jet stream in conjunction with high pressure over coastal Oregon and Washington, is common, even in areas with high annual precipitation. Severe fire events, particularly in moist, high-elevation forests, are usually associated with synoptic weather patterns that drive the interactions among the three air masses (Agee 1993, Ferguson 1997, Schmoldt et al. 1999).

Land use in the ICRB is divided between public (58%) and private (42%) ownership; most forested land is under the jurisdiction of the USDA Forest Service, and the Basin contains 29% of the area covered by wilderness within the contiguous United States (Hann et al. 1997). Conversion of grasslands for agriculture began in the late 19th century, and livestock grazing began in the 1860s. Effective suppression of low- to moderate-severity wildland fires began in the 1930s. Agriculture, grazing, and fire suppression are responsible for major changes in vegetation in both forested and nonforested areas during the last 50–60 yr (Hann et al. 1997).

Geographic databases

The ICRB Landscape Assessment (Hann et al. 1997) and simulation modeling efforts to predict future vegetation (Keane et al. 1996*a*) produced a wealth of GIS coverages, both at a coarse scale (entire ICRB) and a mid-scale (watersheds within national forests [Hessburg et al. 1999]). The coarse-scale coverages provided a geographic template for our model predictions, and a source of predictor variables.

Fire history database.—Our response variable was fire frequency, expressed as the expected (mean) fire return interval (FRI). Although the mean is only one type of measure of central tendency, and the most sensitive to outliers, there are several reasons why we chose it, rather than other measures (e.g., median or mode), as our response variable: 1) many of the fire history studies we used report only the mean (Heyerdahl et al. 1995); 2) the mean is the best way to compare FRIs computed by different methods (for example, a natural fire rotation translates more easily to a mean than to a median or mode); 3) the mean is the most easily integrated into simulation models, where fire frequency is often a random variable with a specified mean and variance (Schmoldt et al. 1999); and 4) there is some evidence that the mean is less sensitive than the other measures to errors from lack of cross-dating tree ring records (D. McKenzie and A. Hessl, *unpublished data*).

From the FHDB, we extracted the following variables for all fire history sites within the ICRB as defined by the maps included in the integrated assessment: 1) fire return interval (response variable: mean = 50.7 yr, range = 6–419 yr), and 2) elevation and geographic coordinates from an Albers projection (predictor variables).

The fire history data vary in quality. Most are not accurately cross-dated, and more than half of the reconstructions use fewer than ten trees. Also, beginning and ending dates vary, admitting possible confounding effects from changing climatic regimes and human activities. The methods employed for calculating fire frequency also differ, and include point estimates, composite fire intervals (CFIs), and natural fire rotation (NFR), or fire cycle computation (Agee 1993, Johnson and Gutsell 1994). The choice of which sites to include in a model involves (1) a trade-off between better model precision, if only the highest quality reconstructions are included (and criteria for inclusion would be partly subjective), and (2) better geographic coverage and representation of high-severity fire regimes, if all reconstructions are included. The confounding effects are likely to be worst for sites experiencing high-severity regimes; these are already underrepresented in the database, and therefore all are important. Also, geographic coverage would be greatly reduced if these sites were eliminated, and we needed as broad a coverage as possible to make predictions for the ICRB.

We therefore decided to develop two model databases; the first would have the advantage of larger sample size and the second the advantage of greater homogeneity: 1) "full data" included as many of the sites as were amenable to minimal standardization, and 2) "reduced data" used a subset of the sites from the first set in which fire frequency had been computed from CFIs (the most common method) and from at least two trees. Different results from modeling these two databases could also be compared as a test of the sensitivity of model behavior to inclusion/exclusion of sites.

The full data database included all 192 sites within the ICRB at which fire frequency had been estimated for areas of <40 ha (Heyerdahl et al. 1995). We employed the following temporal standardization to minimize the confounding effects of climatic variation and, specifically, fire exclusion in the 20th century. For all sites whose histories extended before 1700 and after 1920 (before the period of successful fire suppression in the ICRB), and which included fire dates, we recomputed FRIs based on the years 1700–1920. Sites for which this calculation was impossible (i.e., no fire dates supplied by authors) were dropped. The exception to this was sites with only two fires (e.g., one fire in 1500 and another in 1919). Sites with only one tree

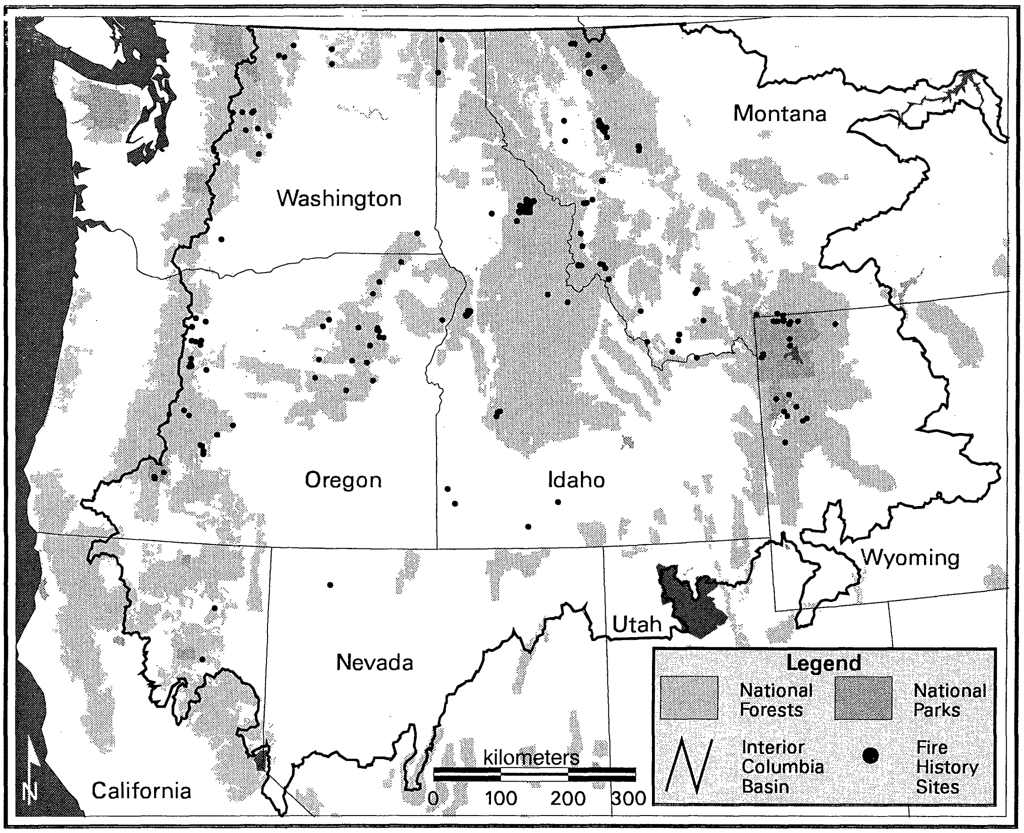

FIG. 1. Fire history sites in the interior Columbia River basin. Many of the sites completely overlap at the resolution of the geographic coordinates.

were retained, but given a smaller importance value in the models (see *Methods: Model development*). Non-cross-dated sites were also retained because they were the majority of sites (157 of 192), but were the basis of a sensitivity analysis of the effects of cross-dating errors on the models (see *Methods: Sensitivity analysis*). After the temporal standardization, the full data contained 185 sites. The reduced data database consisted of all sites from the full data in which FRI had been computed by CFI from at least two trees. Only 90 sites fulfilled this criterion.

We extracted predictor variables from the FHDB for which there are concomitant variables in the ICRB geographic databases (elevation and Albers north and east coordinates), because we were using the model to make predictions for the entire ICRB. We then created a point coverage in ARC-INFO (ESRI 1997) of the fire history site locations in the ICRB, using the Albers projection (Fig. 1).

Other predictors were taken from the following databases:

1) VEG database. Three types of vegetation coverages (ARC-INFO, GRID module, 1-km resolution) were produced for the ICRB coarse-scale simulation effort using satellite imagery and selected comparisons with the midscale coverages derived from photointerpretation and ground truthing (Keane et al. 1996*a*, Hann et al. 1997, Hessburg et al. 1999): 1) historical potential vegetation (PVTH), derived from biophysical parameters including topography, climate, and geomorphology; and 2) historical (~1930); and 3) current dominant cover types (COVH and COVC), derived from the Society of American Foresters classification (Eyre 1980). Each coverage contains both forested and nonforested vegetation types, but we applied model predictions only to the former (Table 1), because the reconstructions in the fire history database were from forested sites only.

2) PRECIP database. Mean annual and summer (June–September) precipitation over the years 1961–1990 (4-km resolution GRID coverage) for the continental United States produced by the PRISM model (Daly et al. 1994). No coverages for the period 1700–1920 are available at this or finer resolutions, therefore we had to assume that spatial patterns of variation in precipitation for 1961–1990 correspond to those from the historical period.

To extract predictors from the VEG database, we overlaid each vegetation coverage with the fire history point coverage, and assigned each fire history site the

TABLE 1. Forested vegetation types in the Interior Columbia River basin.

Potential natural vegetation types	Cover types	Aggregated Küchler types
Interior ponderosa pine	Interior ponderosa pine	Ponderosa pine
Dry Douglas-fir with ponderosa pine	Pacific ponderosa pine	Mixed conifer
Dry Douglas-fir without ponderosa pine	Sierra mixed conifer	Western oakwoods
Pacific pine/Sierra mixed conifer	Oregon white oak	Great Basin pine
Dry grand fir	Limber pine	Lodgepole pine
Oregon white oak	Mixed conifer woodland	Douglas-fir
Limber pine	Douglas-fir	Cedar/hemlock/pine
Moist Douglas-fir	Grand fir/white fir	Silver fir/Douglas-fir
Grand fir E. Cascades	Western larch	Western fir/spruce
Grand fir inland	Lodgepole pine	
Lodgepole pine Oregon	Aspen	
Mountain hemlock/Shasta red fir	Western white pine	
Aspen	Shasta red fir	
Cedar/hemlock E. Cascades	Western hemlock/western red cedar	
Cedar/hemlock inland	Pacific silver fir	
Pacific silver fir	Mountain hemlock	
Mountain hemlock E. Cascades	Engelmann spruce/subalpine fir	
Mountain hemlock inland	Whitebark pine	
Lodgepole pine Yellowstone	Whitebark pine/subalpine larch	
Spruce–fir with aspen		
Spruce–fir without aspen		
Spruce–fir wet		
Spruce–fir WBP > LPP†		
Spruce–fir LPP > WBP‡		

Note: Aggregated Küchler types were determined by combining cover types according to McKenzie et al. (1996*a*).
† More whitebark pine than lodgepole pine.
‡ More lodgepole pine than whitebark pine.

vegetation type of the pixel into which it fell. For the PRECIP database, because the pixel values are discrete approximations of a continuous surface, we used the LATTICESPOT command in GRID to obtain, for each fire history point, a distance-weighted average for the pixel it was in plus the four adjacent pixels.

Quantifying vegetation types

The initial model matrix comprised a mix of qualitative and quantitative data. Vegetation types could be modeled as categorical variables, but this would eliminate the possibility of extrapolating the model to forested sites in the ICRB that had vegetation types not represented in the model database. We developed a qualitative clustering procedure to assign numerical values to them, based on the type of fire regime we expected to be associated with each type. This process was analogous to testing various transformations of predictor variables to identify those with the strongest correlations with the response (Neter et al. 1990).

Within each classification (potential vegetation types [PVTH] and cover types [COVH and COVC]), we ranked the vegetation types initially according to what we expected to be their average FRIs. We also assigned a "distance" between each pair of adjacent types, representing qualitatively the ecological distances, with respect to fire regime, between them. The resulting hierarchical model is represented as a dendrogram (Figs. 2 and 3), an ordered classification of vegetation types that can be viewed at several levels of aggregation (just as, for example, organisms can be viewed at several taxonomic levels). Unlike dendrograms produced from cluster analysis, the order of the leaves is fixed. Beginning at the top of the dendrogram, each type (represented at the leaves of the dendrogram) was assigned a dimensionless numerical value, beginning with 1, based on its distance in the dendrogram from the previous type (Tables 2 and 3). This distance was calculated as the level of aggregation reached in the shortest traverse from one type to the next along branches of the dendrogram. For example, in Fig. 3, the path between "Dry grand fir" and "Oregon white oak" reaches aggregation level 4, thus the numerical value "Oregon white oak," at level 0, is 8 (from the previous value for "Dry grand fir") plus 4 (from the aggregation level reached by the path between the two types) = 12 (Table 2). As the level of aggregation increased (moving from left to right in the dendrogram: Figs. 2 and 3), numerical values became closer together.

At completion of this process, each vegetation type (PVTH, COVH, COVC) had either three or four numerical values associated with it (Tables 3 and 4). We then compared the dendrograms to variants (created by rotating leaves) with respect to correlations (Pearson's R) between the resulting numerical values and the observations of fire frequency in the FHDB. The final dendrograms exhibited the highest correlations at aggregation levels 0 and 1 (Figs. 2 and 3) of any of the variants we examined. We assigned only integer values to each vegetation type, but we explored nonlinear

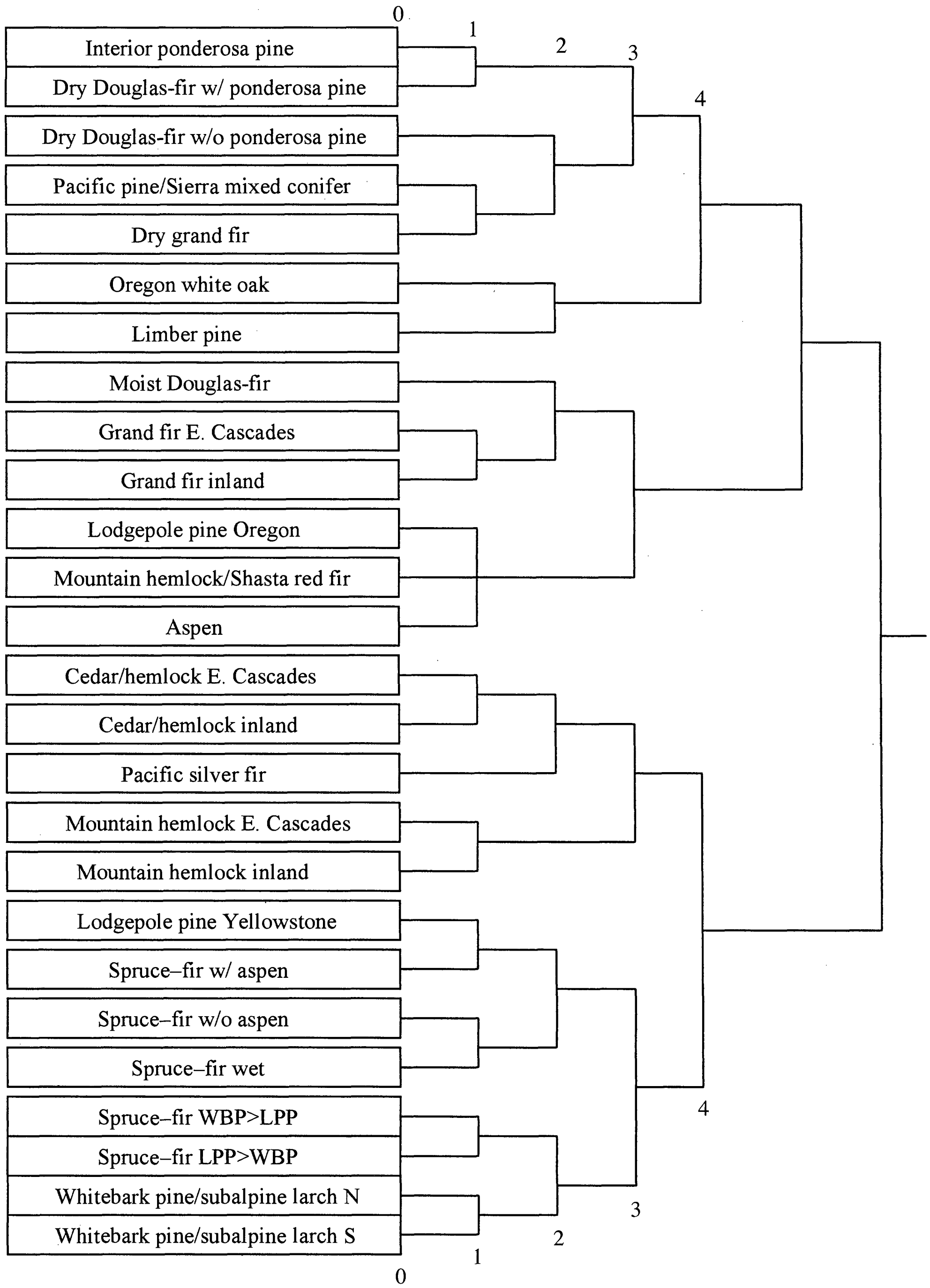

FIG. 2. Dendrogram of forested potential vegetation types (PVTs) in the Interior Columbia River basin, arranged so that proximity in the dendrogram represents similarity in fire regime. Aggregation levels correspond to those in Table 2.

transformations of them during model development to optimize their predictive power.

Model development

We used Splus, version 3.3 for Windows and version 3.4 for UNIX (Mathsoft 1994) for the statistical modeling. Each procedure described below was performed on both the full data and reduced data.

Fire, like many disturbances, is a contagious process (Turner et al. 1989), meaning that fires that affect one location (pixel) are likely to affect nearby pixels. Thus, with enough data points, one would expect some au-

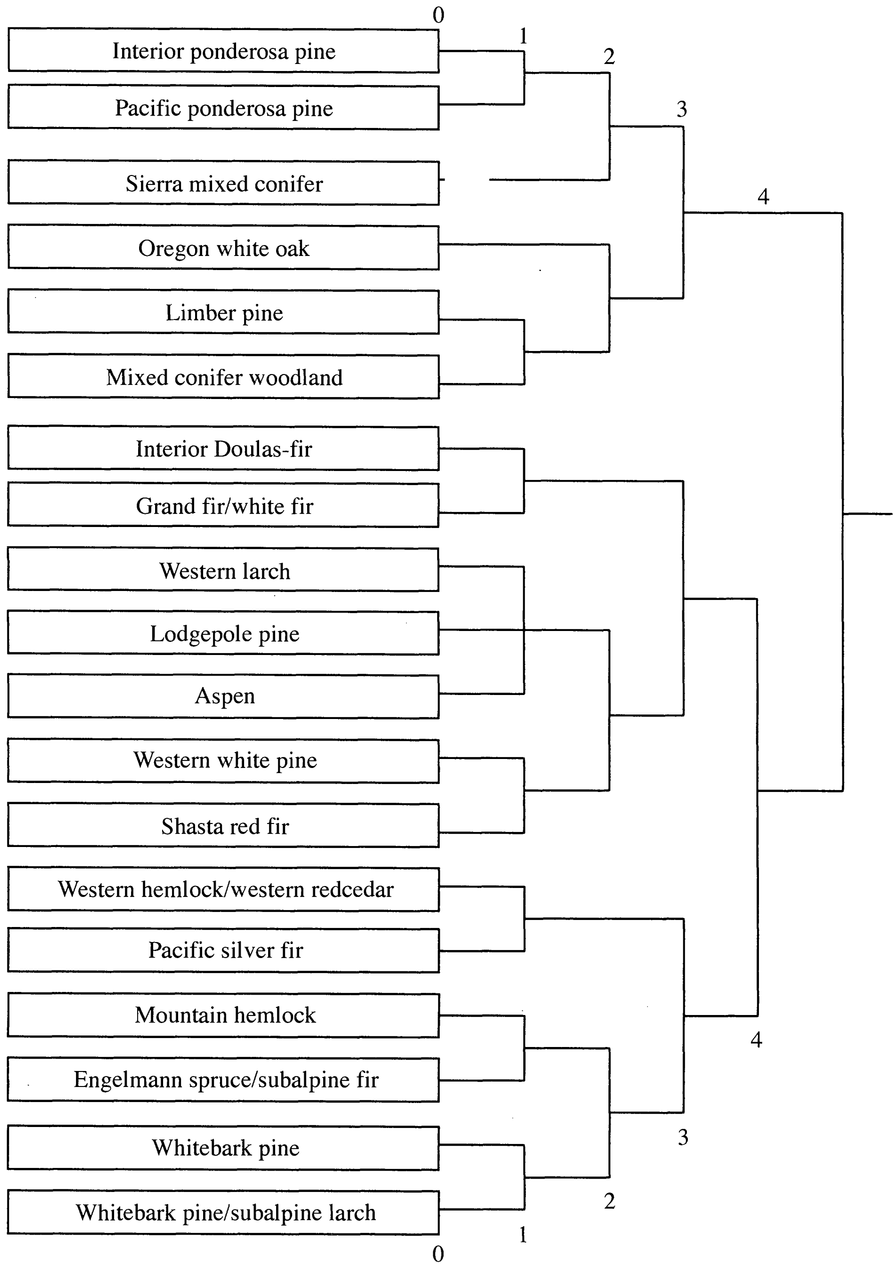

FIG. 3. Dendrogram of forested vegetation cover types (COVs) in the interior Columbia River basin, arranged so that proximity in the dendrogram represents similarity in fire regime. Aggregation levels correspond to those in Table 3.

tocorrelation structure to be evident when examining mean FRIs at fire history sites, depending on the nature of barriers to fire spread between the sites. We examined the spatial covariance among FRIs by computing variograms for clusters of data points (selected visually from Fig. 1), assuming that if spatial dependence was not evident at any scale within the clusters we could assume independence between sites in different clusters. Variogram models could not be parameterized, due to the high variability in FRI between neighboring points. Thus, even though we could not test for spatial independence of the fire regimes represented by the FRIs, we were able to assume statistical independence in the response variable. We then searched for optimal models of two types: 1) a weighted multiple regression of FRI on predictor variables (Neter et al. 1990), and 2) a weighted tree-based (nonparametric) model of FRI on predictor variables (Breiman et al. 1984). In multiple regression, a weighted sum of squares of residuals was minimized, whereas in the tree-based models,

TABLE 2. Numerical values at four levels of aggregation for potential vegetation types, used to create numerical predictors in the model database.

Potential vegetation type	None	One	Two	Three
Interior ponderosa pine	1	1	1	1
Dry Douglas-fir with ponderosa pine	2	1	1	1
Dry Douglas-fir without ponderosa pine	5	3	2	1
Pacific pine/Sierra mixed conifer	7	4	2	1
Dry grand fir	8	4	2	1
Oregon white oak	12	7	4	2
Limber pine	14	8	4	2
Moist Douglas-fir	19	12	7	4
Grand fir E. Cascades	21	13	7	4
Grand fir inland	22	13	7	4
Lodgepole pine Oregon	25	15	8	4
Mountain hemlock/Shasta red fir	26	15	8	4
Aspen	27	15	8	4
Cedar/hemlock E. Cascades	33	20	12	7
Cedar/hemlock inland	34	20	12	7
Pacific silver fir	36	21	12	7
Mountain hemlock E. Cascades	39	23	13	7
Mountain hemlock inland	40	23	13	7
Lodgepole pine Yellowstone	44	27	15	8
Spruce–fir with aspen	45	27	15	8
Spruce–fir without aspen	47	28	15	8
Spruce–fir wet	48	28	15	8
Spruce–fir WBP > LPP†	51	30	16	8
Spruce–fir LPP > WBP‡	52	30	16	8
Whitebark pine/subalpine larch north	54	31	16	8
Whitebark pine/subalpine larch south	55	31	16	8

Note: Based on the dendrogram in Fig. 2.
† More whitebark pine than lodgepole pine.
‡ More lodgepole pine than whitebark pine.

weighted means were calculated at each partition. For both types of models, the response variable was weighted in the following ways:

1) Full data: sites that were cross-dated or had FRIs computed from 10 or more trees were given a weight of 1.0 (37 sites). Remaining sites with more than two trees were given a weight of 0.5 (52 sites). Others were weighted at 0.25 (96 sites).

TABLE 3. Numerical values at three levels of aggregation for cover types, used to create numerical predictors in the model database.

Cover type	None	One	Two
Interior ponderosa pine	1	1	1
Pacific ponderosa pine	2	1	1
Sierra mixed conifer	4	2	1
Oregon white oak	7	4	2
Limber pine	9	5	2
Mixed conifer woodland	10	5	2
Interior Douglas-fir	15	9	5
Grand fir/white fir	16	9	5
Western larch	19	11	6
Lodgepole pine	20	11	6
Aspen	21	11	6
Western white pine	23	12	6
Shasta red fir	24	12	6
Western hemlock/western red cedar	28	15	8
Pacific silver fir	29	15	8
Mountain hemlock	32	17	9
Engelmann spruce/subalpine fir	33	17	9
Whitebark pine	35	18	9
Whitebark pine/subalpine larch	36	18	9

Note: Based on the dendrogram in Fig. 3.

2) Reduced data: sites that were cross-dated or had FRIs computed from 10 or more trees were given a weight of 1.0 (48 sites). All others were weighted at 0.5 (42 sites).

Multiple regression of FRI on predictors in the model matrix.—We developed an exhaustive procedure that tested combinations of the environmental variables (PRECIP, AlbersN, AlbersE) with each set of numerical values (corresponding to a level of aggregation in the dendrogram) for the three ICRB vegetation classifications. We then used backward elimination (Neter et al. 1990) to remove predictors that did not contribute significantly ($P > 0.05$) to the reduction in variance. The response variable was transformed as necessary to meet the normality assumptions of regression, and a Cook's distance plot (Neter et al. 1990) was used to identify and remove significant outliers. Once a model was selected, we compared the output from robust regression to that from ordinary regression.

To find the optimal transformation of the numerical values for vegetation types, we compared a $\log_e$ transform to fitted exponents for the vegetation variables. We used partially linear least squares (Bates and Lindstrom 1986) to obtain the extra coefficient. For example, a possible model form would be

$$\log(\mathrm{FRI}) = \beta_0 + \ldots + \beta_k(\mathrm{COVH})^{\beta_{k+1}} + \ldots \quad (1)$$

where all coefficients are linear except β_{k+1}.

Tree-based model of FRI on predictors in the model matrix.—Tree-based models are a nonparametric alter-

TABLE 4. Parameter estimates for the regression models.

Model	Independent variable	Coefficient (1 SE)	P
Full data ($\log_e$ transform of response variable)	Intercept	−2.1175 (0.8226)	0.0109
	$\log_e$(COVH1)	0.1526 (0.0649)	0.0198
	AlbersN	3.0820×10^{-6} (4.7924×10^{-7})	<0.0001
	Summer precipitation (PPTSUM)	1.2422×10^{-2} (3.2753×10^{-3})	0.0002
	Elevation	2.7546×10^{-3} (5.2565×10^{-4})	<0.0001
	Precipitation/elevation interaction	-9.7262×10^{-6} (2.4182×10^{-6})	<0.0001
Reduced data (square-root transform of response variable)	Intercept	−13.9152 (4.3992)	0.0022
	AlbersN	8.1725×10^{-6} (2.8509×10^{-6})	0.0053
	Summer precipitation (PPTSUM)	0.0814 (0.0168)	<0.0001
	Elevation	9.8024×10^{-3} (2.8242×10^{-3})	0.0008
	Precipitation/elevation interaction	-4.6789×10^{-5} (1.2978×10^{-5}	0.0005

Notes: COVH1 refers to level aggregation "One" in Table 3. The intercept term for the full data is corrected for logarithmic bias (Flewelling and Pienaar 1981).

native to linear models for regression problems (Breiman et al. 1984). They are fit by binary recursive partitioning, in which a data set is successively split into increasingly homogeneous subsets, using a likelihood criterion to maximize the reduction in deviance produced by each partition (Clark and Pregibon 1992). Although they are often used as an exploratory technique for revealing structure in data, they can also be used for prediction when predictors in a new database fall within the range of predictors in the modeling database. A particular advantage of tree-based models is that they can capture nonadditive behavior and complex interactions between variables, whereas standard linear models are limited to prespecified multiplicative interactions (Clark and Pregibon 1992). Response variables that are factors produce classification trees, whereas numerical response variables produce regression trees.

Our tree-based models were built from the same model databases as the regression models. We used an adaptive estimation method (Breiman et al. 1984) to minimize the complexity of the model (number of branches and nodes) without sacrificing goodness of fit. We first fit an overly large tree, using two criteria for deciding when a node should not be split: 1) if node deviance is <1% of the root node deviance, or 2) if the node has fewer than 10 observations. We then used a cost-complexity measure derived by Breiman et al. (1984) to prune the tree:

$$\mathrm{D}_\alpha(T_i) = \mathrm{D}(T_i) + \alpha(\mathrm{size}(T_i)) \qquad (2)$$

where

$$\mathrm{D}(T_i) = \text{deviance of subtree } T_i$$

$$\mathrm{size}(T_i) = \text{the number of terminal nodes of } T_i$$

$$\alpha = \text{a cost-complexity parameter.}$$

For a specified α, the cost-complexity pruning implemented in Splus minimizes $\mathrm{D}_\alpha(T_i)$ for all subtrees of a tree T. We determined α graphically by plotting deviance against number of nodes as a step function. The value of the cost-complexity parameter corresponding to the flattening of this step function was inserted in Eq. 2. Each subtree was pruned, beginning at its terminal nodes, until the measured cost complexity was minimized. Predictor variables on which there were no partitions in the final pruned model were thus eliminated.

Large-scale application of the models.—For each variable in the final (tree-based or regression) models, we created a raster coverage (GRID: 1-km resolution) with data values only at forested pixels. A pixel was considered forested if both its corresponding pixels in the coverages of historical potential vegetation and of dominant cover type coincided with the vegetation types in Table 1. We exported these GRIDs to Splus and used the tree-based and linear regression models to predict the FRIs for the new data (all forested pixels within the ICRB). We then imported the predicted values into ARC-INFO, creating four raster coverages of predicted FRIs. Because tree-based models are inaccurate when extrapolated beyond the range of model databases, we eliminated pixels from the tree-based GRID that corresponded to values of environmental

TABLE 5. Cover types (COVH) in the (complete) fire history database vs. the ICRB coarse-scale vegetation coverage.

Cover type	Number of sites in fire history database	Number of pixels in ICRB (1 km) coverage	Percentage of total forested pixels
Interior ponderosa pine	65	97 762	31.2
Pacific ponderosa pine	0	2539	0.8
Sierra mixed conifer	0	872	0.3
Oregon white oak	0	481	0.2
Limber pine	0	263	0.1
Interior Douglas-fir	12	49 786	15.9
Grand fir/white fir	7	4210	1.3
Western larch	32	21 338	6.8
Lodgepole pine	42	67 347	21.5
Aspen	6	8888	2.8
Western white pine	7	10 477	3.3
Shasta red fir	1	10	<0.1
Western hemlock/western red cedar	1	404	0.1
Pacific silver fir	0	123	<0.1
Mountain hemlock	0	824	0.3
Engelmann spruce/subalpine fir	19	30 655	9.8
Whitebark pine	0	15 120	4.8
Whitebark pine/subalpine larch	0	2108	0.7
Total	192	313 207	100.0

variables outside those in the fire history databases (elevation > 2550 m and annual precipitation > 2000 mm = 9% of total area excluded). We did not eliminate any pixels based on vegetation type, however, because all were within the range of numerical values of predictors from the FHDB (Tables 3 and 5).

Model evaluation

The evaluation of any statistical model typically addresses the following questions:

1) How well does the model fit the data and meet assumptions?

2) How well does the model predict new observations?

We addressed Question 1 with standard diagnostics, including plots of residuals against fitted values to identify heterogeneous variance, and quantile–quantile plots to assess normality of residuals (Neter et al. 1990). Because the model databases were small and very heterogeneous, we did not create a subset of the database for testing. Instead, to answer Question 2, we calculated a bootstrap estimate of prediction error (Efron and Tibshirani 1993) for the regression models, and compared it to the model's error sum of squares (SSE). For the tree-based models we used 10-fold cross validation to assess graphically the degree of pruning that we applied to the model tree (Venables and Ripley 1994). However, because the purpose of our model was to extrapolate local relationships to the regional scale, there were two other questions we needed to address:

3) How well does the model database (point information) represent the entire region?

4) How are model behavior and concomitant errors propagated in the extrapolation process?

To answer Question 3, we compared the distributions of predictor variables in the model database to those in the regional database. For example, predictions for a cover type that was abundantly represented in the region, but only sparsely represented in, or absent from, the model database, could be suspect. Conversely, predictions for a pixel whose environmental variables were well within the range of the model database, and whose cover type was abundantly represented therein, could be accepted with more confidence. This procedure also suggested which types of sites were likely candidates for future fire history studies by virtue of underrepresentation in the model database.

To answer questions 2 (at larger scales) and 4, we produced statistical and graphical summaries of model predictions at the 1-km scale, using two different levels of aggregation of vegetation types: historical cover types and aggregated Küchler types (McKenzie et al. 1996*c* and Table 1). We examined the distribution of predicted FRIs from both models for each vegetation type for obvious anomalies, using the output maps and histograms of FRIs for each vegetation type. For example, we expected FRIs to be positively correlated with latitude and elevation, and to observe differences between types in mean and range. We also expected that most predicted FRI distributions for vegetation types would not display major discontinuities or distinctly multimodal patterns. This partly qualitative procedure suggested which of the models would be robust to extrapolation.

Sensitivity analysis.—We expected the principal source of error in both the full and reduced data sets to be the lack of accurate cross-dating for many of the reconstructions that used fire scars. Dendrochronological cross-dating greatly increases the probability that tree ring records will be synchronized so that missing or false rings will not distort the association of distinct events (e.g., as recorded by fire scars) with specific years (Fritts and Swetnam 1986). Drought-sensitive tree species are especially likely to have missing rings (Dieterich and Swetnam 1984). In systems with short FRIs, fire scar dates estimated by ring counting and matching samples may differ from dates estimated by cross-dating (Madany et al. 1982). Not only could significant errors occur without cross-dating, but also consistent biases could arise from the tendency of a particular researcher to be a "lumper" or a "splitter." For example, lumpers might assume that fire scars that appeared to be one or more years apart were actually from the same fire, whether or not this was the case. Conversely, splitters might assume, trusting their ring counts, that fire scars even one year apart always represented different fires.

To estimate the magnitude of errors, we simulated increment cores with fire scars by creating time series of random fire intervals. From rough calculations on these, and discussions with fire ecologists, we concluded that a typical error would be to underestimate fire frequency by a factor of two. Less commonly, one might overestimate by a factor of two. A simple example of this would be two samples whose fire dates appeared to be consistently out of synchrony by one or two years, leading a lumper to assume that each pair represented only one fire. To estimate the effects of this and similar errors on the parameters and broad-scale behavior of the models, we simulated a random correction factor that could be applied to noncross-dated FRI estimates in the fire history database to account for potential cross-dating errors. The factor was applied differently in two scenarios:

1) Lumpers. This scenario assumes that estimates of FRI will be high because researchers would tend to adjust fire dates from different samples to be more synchronized. The correction adjusted FRIs down, on average, to half of their observed values.

2) Lumpers and splitters. This scenario assumes that errors will be equally likely on either side of the original. Corrected FRIs were higher (twice the observed) or lower (half the observed), on average, with equal probability.

Because the correction factor is a random variable in both scenarios, every correction is unique, and every realization of a set of corrections applied to the FRIs is also unique. We applied each scenario 25 times to both regression models and both tree-based models, correcting only FRIs under 30 yr from noncross-dated fire scars (Heyerdahl et al. 1995), and storing the parameter estimates and their p values, the fitted values, and R^2 (regression models) and proportional reduction in deviance (PRD) (tree-based models). We compared these data to output from the original four models, and randomly selected realizations (of regression models only) to compare to model predictions at the regional scale. We were particularly interested in the following responses:

1) Sensitivity of parameter estimates (regression models) or order of partitions (tree-based models).

2) Major changes in the significance levels of parameter estimates (regression models) or deviance reductions at splits on a particular parameter (tree-based models).

3) Proportional changes in predicted values at the regional scale.

RESULTS

For the full data, the best multiple regression model uses four predictor variables (Table 4), is highly significant (n = 182, $P < 0.0001$), and has reasonable explanatory power ($R^2 = 0.44$). The model for the reduced data uses three predictors (Table 4), is also highly significant (n = 87, $P < 0.0001$), and has better explanatory power ($R^2 = 0.57$). In both models, three outliers were removed based on Cook's distance. Signs of coefficients, except for interaction terms, are positive, thus an increase in summer precipitation, latitude, elevation, or the numerical value of COVH1 increases predicted FRI. Standard diagnostic procedures revealed no violation of regression assumptions in either model, and coefficients from the robust procedure are virtually identical to those from ordinary regression. The range of fitted values for FRI is 8–87 yr for the full data, and 3–124 yr for the reduced data. The bootstrap estimate of prediction error (from 100 replicates) produced 6.5% and 7.3% error inflation above SSE for the full data and reduced data models, respectively. The correlation (Pearson's R) between fitted values for the two regression models (on sites common to both) was 0.95.

The tree-based model for the full data, after pruning, produced 16 distinct predicted values, ranging from 8 to 131 yr (Fig. 4a). The model uses the same four variables as the full data regression model (minus the interaction term; Table 4). The primary partition is on AlbersN (latitude), which accounted for 49% of the total reduction in deviance. The number of sites represented by terminal nodes ranges from 5 to 26. For the reduced data, there are only 10 distinct predicted values, ranging from 11–150 yr (Fig. 4b). The primary partition is also on AlbersN, accounting for 70% of the total reduction in deviance. The number of sites represented by terminal nodes ranges from 5 to 25. PRD from both tree-based models (roughly equivalent to R^2) is 0.77; hence, they have greater explanatory power, in the statistical sense, than the regression models. Cross validation indicated that more severe pruning might also be acceptable, but we wanted to retain as broad a range, and as great a variety, of fitted values as possible because of the number of predictions we were making from the models. Therefore we retained all the nodes remaining after the cost-complexity pruning.

Sensitivity analysis: model details

Parameter estimates in the regression models changed little (maximum change less than 1% for any

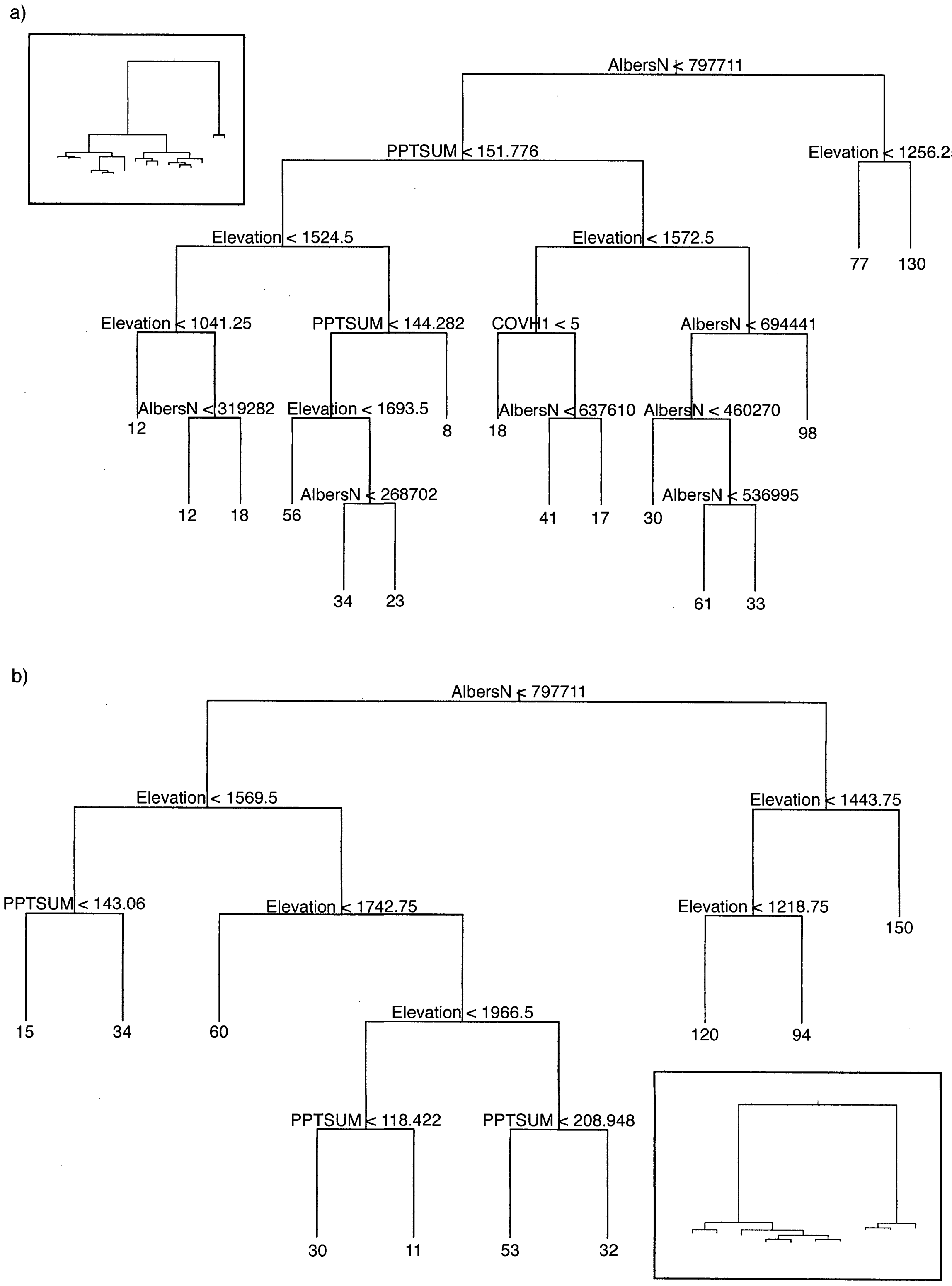

FIG. 4. The final regression trees for (a) full data and (b) reduced data. Values at nodes are the predicted fire return intervals, rounded to whole years. Significant digits reflect the resolution of the predictors. Predictions for a new site are obtained by moving down the tree, branching left at a split if the site meets the rule, and branching right otherwise. In the insets, trees are scaled to display the proportion of variance accounted for at each split. COVH1 refers to the first level of aggregation of COVH (Table 3).

parameter) in either Scenario 1 or Scenario 2. For the full data model, the highest fitted values from Scenario 1 were slightly lower on average (10%) than for the "true" model. For the reduced data model, the lowest fitted values from Scenario 2 were 50% lower on average than for the true model. Other extrema of fitted values differed <1% from the true models.

For the full data, model R^2 was consistently lower by 12–15% in Scenario 1, and by 3–6% in Scenario 2. For the reduced data, R^2 was similar to that of the true model, differing by <4%, for both scenarios. The parameter COVH1 (full data only) lost significance, however, in 96% of the runs of Scenario 1 and 76% of the runs of Scenario 2, indicating that this parameter may not be very robust to cross-dating errors.

The tree-based models changed very little in either scenario. Primary partitions remained on AlbersN, PRDs changed only 1–5%, and no major structural changes occurred. Fitted values at terminal nodes changed <10%, and splits on the predictors were consistent. For example, the partition of AlbersN at 797711 (Fig. 4) in both true models was retained through all iterations of both scenarios.

Model behavior

The total number of regional-scale predictions from the models is three orders of magnitude greater than the number of sample sites (Table 5). The regional predictions cover a larger elevational range (49–3713 m) than the model database (727–2550 m). Because COVH1 was the vegetation variable represented in both models, we used COVH types to organize summary statistics for the models. There are eight COVH types in the regional (forested) coverage that were not represented in the model database, although these account for <8% of the total pixels. Predictions of FRI from the regression models range from 1 to 375 yr at the regional scale for the full data model, and 2 to 290 yr for the reduced data model. Predictions from the tree-based models are restricted to the 16 (full data) and 10 (reduced data) discrete values at the nodes of the respective trees.

Viewed regionally, predictions from the regression models reveal latitudinal gradients (Fig. 5). The gradient is the dominant feature of the reduced data model (Fig. 5a); the full data model predicts that the longest FRIs will be in the northern Cascade Mountains, Washington, the Wind River Mountains, Wyoming, and in the northwestern corner of Montana (Fig. 5b). Predictions from the tree-based models display distinct horizontal bands in addition to the latitudinal gradient (Fig. 6). These bands do not correspond to known biotic or abiotic gradients and are artifacts of the dominance of AlbersN in the partitioning process and of the limited number of unique predicted values (16 and 10). When separated by COVH type, most distributions of predictions from the regression model are unimodal and right skewed, whereas predictions from the tree-based models, particularly the reduced data model, are distinctly bimodal, often with wide separations between modes.

Sensitivity analysis: model behavior

At the regional scale, the predicted FRIs from realizations in the sensitivity analysis (RSAs) closely track those from the corresponding regression model. Except for a few extreme outliers (<0.1% of pixels), differences (true model: RSA) are <10 yr for all comparisons. Proportional differences are much greater in cover types predicted to have short FRIs. For example, for the full data, Scenario 1, the difference is between 1 and 7 yr on 95% of pixels that are "Interior ponderosa pine," and between −5 and 14 yr on 95% of pixels that are "Engelmann spruce/subalpine fir." For the reduced data, Scenario 2, the corresponding differences are between −3 and 1 yr for "Interior ponderosa pine" and between −6 and 2 yr for "Engelmann spruce/subalpine fir." As expected, RSAs from Scenario 1 consistently predict shorter FRIs in ponderosa pine systems (32% of total pixels) than the regression models, because in the lumpers scenario, FRIs for these sites in the fire history database were assumed to have been overestimated. However, this consistent bias is not apparent for "Interior Douglas-fir," the other common vegetation type for which many FRIs were reduced in the RSAs, Scenario 1.

Discussion

The models reveal highly significant relationships between fire frequency and the predictor variables. The data represented by the output maps provide new information, which will complement the ICRB assessment and assist coarse-scale modeling efforts in the region. In contrast to the coverages from the ICRB assessment, which delineate five broad ranges of fire frequency (Morgan et al. 1996), our models produce estimates of fire frequency at the resolution of one year. Also in contrast, the ICRB models assigned fire regime classes to cover types (Hann et al. 1997), whereas our models predict fire frequency principally from environmental and geographic variables.

We expected the regression model to predict increasing FRIs along both elevational (low to high) and latitudinal (south to north) gradients, and a significant positive correlation between FRI and PPTSUM, given the obvious link between fuel moisture and flammability and extensive documentation of longer FRIs in more mesic systems. The resolution of the PRISM coverage (4 km) may be too coarse to capture fine-scale fluctuations in precipitation that could have significant differential effects on fuels and on fire ignition and spread. Precipitation maps with finer resolution might show an even stronger relationship with FRI. For example, historical reconstructions of climate from tree ring data (e.g., Fritts et al. 1979), collected at nearby sites, could provide better predictors. Gridded dendro-

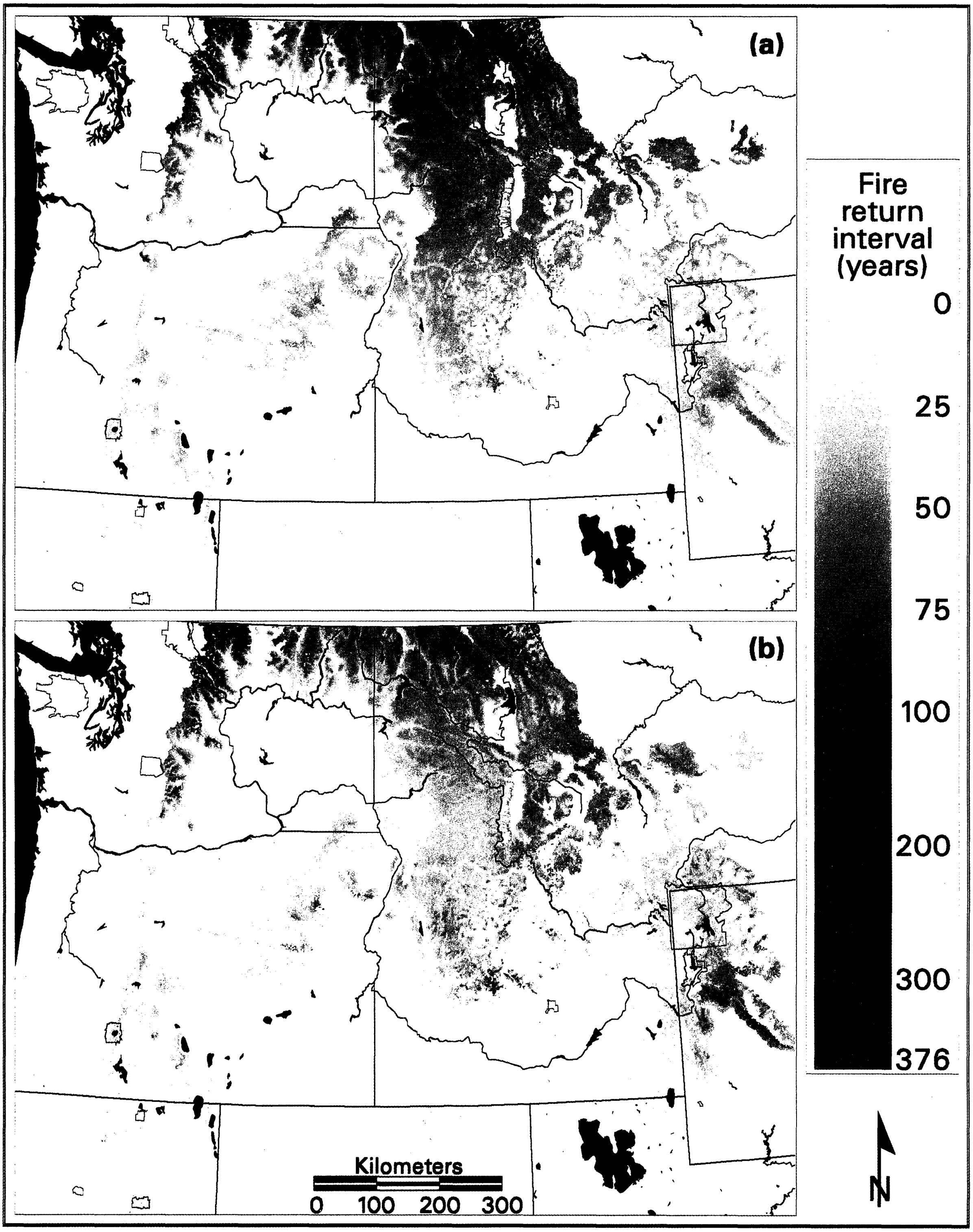

FIG. 5. Output maps of predicted fire return intervals from the regression models, displayed as a continuous gradation of color over the predicted range for (a) reduced data and (b) full data.

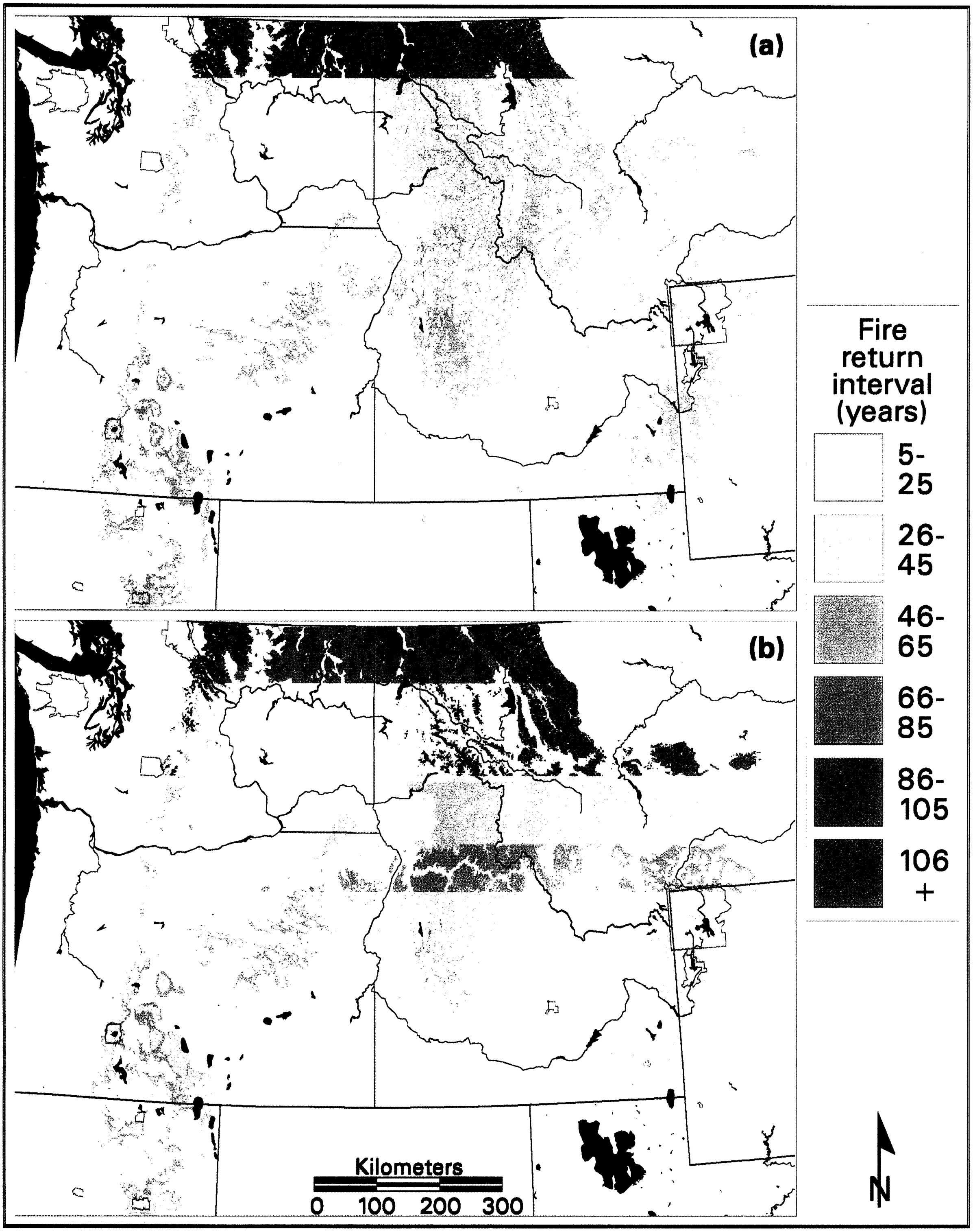

FIG. 6. Output maps of predicted fire return intervals from the tree-based models, collected into six categories. Red shading shows the proportional area of the coverage in each category. (a) Reduced data: 5–25 yr = 30%, 26–45 yr = 33%, 46–65 yr = 24%, 66–85 yr = 0%, 86–105 yr = 2%, >105 yr = 11%. (b) Full data: 5–25 yr = 34%, 26–45 yr = 32%, 46–65 yr = 14%, 66–85 yr = 6%, 86–105 yr = 7%, >105 yr = 7%.

climatic reconstructions (Cook et al. 1999) could then be used for regional-scale predictions.

The negative coefficient for the elevation/precipitation interaction in both regression models indicates that at higher elevations, FRI is less strongly correlated with precipitation than at low elevations. For example, the models predict that the differences in FRI between low-elevation ponderosa pine forests (drier) and low-elevation cedar-hemlock forests (wetter) would be proportionally greater than between high-elevation whitebark pine forests (drier) and high-elevation mountain hemlock forests (wetter). We expected that of the three vegetation variables, COVH would be correlated most strongly with FRI, because it represents vegetation presumed to be present during the period 1700–1920. It also reflects the dominant successional process (fire). COVH1 is the best predictor of the vegetation types (although it only appears in the full data model), reinforcing the idea that fire regimes in the ICRB could be classified by aggregated vegetation types (see also Morgan et al. 1996).

The tree-based and regression models serve complementary purposes in understanding the relationships among FRI and the predictor variables. Results of the tree-based models suggest that more variation in the response can be explained by exploring complex interactions and dependencies among variables than can be captured by an ordinary regression model (PRD = 0.77 vs. R^2 = 0.44 and 0.57). However, when extrapolated to 1-km resolution, the tree-based models produce horizontal bands of the same predicted values and bimodal distributions of FRIs. These artifacts are due to the sequential nature of the prediction process; once a node has been passed, the choices are limited to those further down the branch. The inability of tree-based models to predict new values is also a significant drawback in extrapolations of this magnitude. Conversely, the regression models, although they have weaker (statistical) explanatory power at the scale of the model database, provide a simple and robust method of prediction at the regional scale. Thus, although the tree-based models show a better statistical fit, only the regression models are suitable for broad-scale predictions.

Of the two regression models, the reduced data model has a more homogeneous response variable, uses only abiotic variables as predictors, and produces more homogeneous predictions at the regional scale (dominated by a broad latitudinal gradient; Fig. 5a). The full data model incorporates vegetation, albeit weakly, and isolates geographic areas of long FRIs independent of the latitudinal gradient (Fig. 5b). Sensitivity analyses suggest that errors in computing the response variable would be slightly greater for the full data, and proportionally greater in systems with short FRIs. In the absence of data to systematically test the regional-scale predictions, we suggest that predictions from the reduced data model are probably more accurate in systems with short FRIs, because the percentage of non-cross-dated sites is lower. Long FRIs are underrepresented in the full database, but even more so in the reduced database. For example, of 19 sites with cover type "Engelmann spruce/subalpine fir," only one was retained for the reduced data model. Therefore we expect that estimates of FRI for low fire frequency systems will be better from the full data model. The ICRB model does not appear to produce any latitudinal gradient, probably because it is focused on vegetation types and broad classes of fire frequency. Our full data model displays both the broad gradient and isolated areas of high FRIs, while our reduced data model displays only environmental/geographic gradients.

Aggregation error, spatial heterogeneity, and the reliability of the models

Applying the tree-based models at coarse scales introduces a common form of error in data aggregation. When relationships among variables are nonlinear, characterization of data by simple means will produce consistent errors when relationships are extrapolated across scales (O'Neill 1979, King et al. 1991, Rastetter et al. 1992, Cale 1995, O'Neill 1998). The tree-partitioning process equates predicted values of the response to its mean over increasingly homogeneous subsets of the predictors (Clark and Pregibon 1992). This retains features of the raw data at the scale of modeling, but the discontinuities are magnified such that the proportion of extreme values is exaggerated at the scale of prediction because "mistakes" at any split are propagated down through the tree. The regression models provide a rougher approximation of patterns in the original data, but as linear functions, are more robust to aggregation error. The distributions of predictions from these models will have errors associated with them, but their smoothness suggests a lack of interference from model artifacts in the extrapolation process.

Patterns of spatial heterogeneity affect the connectivity of landscapes with respect to fire, and thus introduce biases into estimates of fire frequency (Lertzman et al. 1998). The 1-km vegetation classifications and digital elevation model mask considerable spatial heterogeneity in vegetation and topography. Applying the models at this scale implicitly produces a constant FRI over 1 km^2, but the patterns of landforms and vegetation within the pixel exert considerable influence on fire severity and spread, resultant fire size, and thus expected FRI (Agee 1998). For example, local cold-air drainages can favor narrow corridors of vegetation characteristic of higher elevations in a matrix of drier, low-elevation vegetation (Agee et al. 1990). Depending on the connectivity of the landscape, FRIs in the corridors and matrix in this context may be identical, or at least more similar than could be expected if they were spatially disjunct.

The models could be improved by additional fire history information for the ICRB, particularly if data collection and interval estimation were standardized, providing better confidence to FRI estimates in model databases. Cross-dating all tree ring records would significantly improve the accuracy of the response variable. Although our sensitivity analysis covered two broad categories of cross-dating errors, there are many possibilities, and it is difficult to know in advance what type of errors are likely in a given reconstruction. Some fire history studies should be initiated specifically to improve regional scale models, and to test predictions of models like ours. A sparse grid of fire history sites could include more vegetation types and be amenable to rigorous quantitative methods of spatial aggregation (Dutilleul 1993, Legendre 1993, Rossi et al. 1993). Future efforts should be concentrated in systems that are currently underrepresented in the model database vs. the regional database, namely high-elevation forests (above 2550 m) supporting low frequency/high severity fire regimes (dominant species are whitebark pine, subalpine larch, mountain hemlock, Pacific silver fir; Table 5).

Model applications

Only the regression models should be applied at broad scales, because they are evidently robust to extrapolation errors. We envision three applications for the regression models, while recognizing that they need to be continually revised as more fire history data are made available. Predictions for individual sites should also be viewed with caution, or given confidence intervals of 5–10 yr on either side, because of the uneven quality of the FHDB.

First, the models provide estimates of FRIs at one-year resolution where only estimated ranges existed previously. Most pixels on the output map have no associated fire history data. Although the limits imposed by the coarse resolution mean that the models will be less useful for fine-scale than for broad-scale applications, local managers may be able to integrate model predictions with local qualitative data and knowledge about systems similar to theirs to better estimate the historical range of variability of fire regimes (Morgan et al. 1994, Landres et al. 1999, Swetnam et al. 1999). For example, in forests that historically experienced high frequency, low severity fires, maps of historical FRIs provide input for decadal-scale planning that includes prescribed burns and complementary silvicultural treatments. Similarly, in forests that experienced low frequency, high severity fires, the maps, in combination with records of fire sizes, can suggest the minimum dynamic area (Pickett and Thompson 1978) and appropriate temporal scales for management plans (Hobbs 1998). In heterogeneous terrain where the 1-km scale predictions are too coarse to be of direct use, understanding the importance of environmental gradients and their interactions for fire frequency may help explain observed differences in fire frequency at finer scales. Model predictions can also be added to broader ecological inventories designed to assist ecosystem management (Keane et al. 1996*b*).

Second, the models provide new data to initialize spatially explicit, coarse-scale simulation models of fire behavior, fire effects, and succession. In mechanistic models (e.g., Keane et al. 1996*c*), predicted FRI values at each pixel could be used directly or taken as means of candidate distributions such as the Weibull (Johnson and Gutsell 1994) from which input FRIs could be chosen randomly. In cell-based models of disturbance on abstract landscapes (e.g., Turner et al. 1989, Turner and Romme 1994), patterns of FRIs taken from the regional map could be incorporated into measures of landscape connectivity and of how that connectivity changes over time.

Third, our approach demonstrates a methodology for integrating existing data and making coarse-scale predictions that is relatively robust to aggregation error. Coarse-scale modeling will probably need to incorporate semiqualitative elements for the foreseeable future (Keane and Long 1998, McKenzie 1998). Our results suggest that heuristic, knowledge-based methods (quantifying vegetation types) and rigorous statistical methods can be successfully combined.

Finally, the tree-based models suggest that a quantitative understanding of how variables interact differently in different parts of their ranges should improve the explanatory power of models. They can also generate hypotheses about how different combinations of variables affect fire frequency in different systems. For example, are factors associated with latitude (e.g., growing season length) more significant for understanding differences in fire frequency at certain elevations and under certain climatic regimes? More generally, are there threshold values of variables across which their effects on fire change rapidly? Understanding such discontinuous behavior in ecological systems will undoubtedly improve our ability to make predictions about their future states (O'Neill 1979) and the ability of fire managers to anticipate changes in fire effects and the consequences for forest ecosystems.

Future directions

The success of qualitative methods depends on the robustness of a knowledge-based approach (Schmoldt and Rauscher 1995). Keane et al. (1996*a*) used a set of transition rules, based on expert knowledge, to model successional pathways in a regional-scale simulation model. Similarly, we used a qualitative clustering method to approximate the numerical contribution of vegetation types to estimates of FRI (Figs. 3 and 4). Improvements in vegetation databases, however, might allow approaches like ours to be fully quantitative. Vegetation classifications are frequently subjective or based on broad qualitative rules (Holdridge 1947, Küchler 1964, Eyre 1980, Bailey 1996), but more empirical classifications are possible (Hargrove and Luxmoore 1998). These can be expressed probabilistically, in terms of fuzzy set membership (Roberts 1996). Doing so could alleviate the worst aspects of the aggregation problem. For example, a pixel that is classified as "Pacific silver fir" at the 1-km scale would be expected to have a different fire history if it were composed of 100% Pacific silver fir vs. 60% Pacific silver fir and 40% Douglas-fir (Agee et al. 1990).

Landscape heterogeneity, including topography and the patchiness of vegetation and fuels, constrains fire sizes and therefore the expected extent of spatial autocorrelation of FRIs (Turner and Romme 1994, Agee 1998, Lertzman et al. 1998, Miller and Urban 1999). We were unable to discern spatial autocorrelation among the sites in the existing fire history database, but new sampling designs for fire history reconstructions might address this problem. For example, if grids were established to measure point FRIs in different systems, autocorrelation structure could be more easily determined, and interpolated values could be compared to predictions from a model that assumes independence, providing a check on the reliability of models like ours.

Ecosystem management is being applied under hierarchical frameworks at multiple spatial scales (Quigley et al. 1996, Sierra Nevada Ecosystem Project 1996, Johnson et al. 1999). For example, the ICRB Ecosystem Management Project is modeling forest succession and disturbance at the regional scale, using semiquantitative methods, and at the watershed scale, using mechanistic models (Haynes et al. 1996). Informed decisions are needed at increasingly broad spatial scales, but in most cases, detailed quantitative data are not available. Integration of existing databases, complementary use of qualitative and quantitative methods, resolution of scale incompatibilities in spatial data (Quattrochi and Goodchild 1997), and more efficient approaches to data collection will improve our understanding of broad-scale interactions among fire, vegetation, and the physical environment.

Acknowledgments

We thank Paul Alaback, Charles Halpern, Peter Landres, Thomas Swetnam, John Vande Castle, and three anonymous reviewers for comments that improved the manuscript. Robert Norheim created the GIS maps. Ernesto Alvarado, Christopher Daly, Emily Heyerdahl, and Miriam Rorig provided GIS data layers. Discussions with Richy Harrod, Amy Hessl, Emily Heyerdahl, and Ken Lertzman helped to refine the sensitivity analysis. Research was funded by the USDA Forest Service, Pacific Northwest Research Station, and the Global Change Research Program of the U.S. Geological Survey, Biological Resources Division.

Literature Cited

Agee, J. K. 1993. Fire ecology of Pacific Northwest forests. Island Press, Washington DC, USA.

Agee, J. K. 1994. Fire and weather disturbances in terrestrial ecosystems of the eastern Cascades. USDA Forest Service, General Technical Report PNW-320, Pacific Northwest Research Station, Portland, Oregon, USA.

Agee, J. K. 1998. The landscape ecology of western fire regimes. Northwest Science (special issue) **72**:24–34.

Agee, J. K., M. Finney, and R. de Gouvenain. 1990. Forest fire history of Desolation Peak, Washington. Canadian Journal of Forest Research **20**:350–356.

Bailey, R. G. 1996. Ecosystem geography. Springer-Verlag, New York, New York, USA.

Baker, W. L. 1989. Effect of scale and spatial heterogeneity on fire-interval distributions. Canadian Journal of Forest Research **19**:700–706.

Baker, W. L. 1995. Long-term response of disturbance landscapes to human intervention and global change. Landscape Ecology **10**:143–159.

Bates, D. M., and M. J. Lindstrom. 1986. Pages 152–157 *in* M. J. Lindstrom, editor. Nonlinear least squares with conditionally linear parameters. American Statistical Association, Proceedings of the Statistical Computing Section, Madison, Wisconsin, USA.

Boychuck, D., A. J. Perera, M. Ter-Mikaelian, D. L. Martell, and C. Li. 1997. Modelling the effect of spatial scale and correlated fire disturbances on forest age distribution. Ecological Modelling **95**:145–164.

Breiman, L., J. H. Friedman, R. Olshen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth International Group, Belmont, California, USA.

Cale, W. G. 1995. Model aggregation: ecological perspectives. Pages 230–241 *in* B. C. Patten and S. E. Jørgensen, editors. Complex ecology: the part–whole relation in ecosystems. Prentice Hall, Englewood Cliffs, New Jersey, USA.

Cattelino, P. J., I. R. Noble, R. O. Slatyer, and S. R. Kessell. 1979. Predicting the multiple pathways of plant succession. Environmental Management **3**:41–50.

Clark, J. 1996. Testing disturbance theory with long-term data: alternative life-history solutions to the distribution of events. American Naturalist **148**:976–996.

Clark, L. A., and D. Pregibon. 1992. Tree-based models. Pages 377–419 *in* J. M. Chambers and T. J. Hastie, editors. Statistical Models in S. Wadsworth and Brooks, Pacific Grove, California, USA.

Cook, E., D. Meko, D. Stahle, and M. Cleaveland. 1999. Drought reconstructions for the continental United States. Journal of Climate **12**:1145–1162.

Daly, C., R. P. Neilson, and D. L. Phillips. 1994. A statistical–topographical model for estimating climatological precipitation over mountainous terrain. Journal of Applied Meteorology **33**:140–158.

Dieterich, J. H., and T. W. Swetnam. 1984. Dendrochronology of a fire scarred ponderosa pine. Forest Science **30**: 238–247.

Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. Ecology **74**:1646–1658.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York, New York, USA.

ESRI (Environmental Systems Research Institute). 1997. ARC/INFO. Version 7. 0. Environmental Systems Research Institute, Redlands, California, USA.

Eyre, F. H. 1980. Forest cover types of the United States and Canada. Society of American Foresters, Washington, DC, USA.

Ferguson, S. A. 1997. A climate-change scenario for the Columbia River Basin. USDA Forest Service Research Paper PNW-499. Pacific Northwest Research Station, Portland, Oregon, USA.

Finney, M. 1998. FARSITE: Fire Area Simulator: model development and application. USDA Forest Service Research Paper RMRS-RP-4, Rocky Mountain Research Station, Ogden, Utah, USA.

Flewelling, J. W., and L. V. Pienaar. 1981. Multiplicative regression with lognormal errors. Forest Science **27**:281–289.

Fosberg, M. A., L. O. Mearns, and C. Price. 1992. Climate change–fire interactions at the global scale: predictions and limitations of methods. Pages 123–137 *in* P. J. Crutzen and J. G. Goldammer, editors. Fire in the environment. Wiley and Sons, New York, New York, USA.

Frelich, L. F., and P. B. Reich. 1995. Spatial patterns and succession in a Minnesota southern boreal forest. Ecological Monographs **65**:325–346.

Fritts, H. C., G. R. Lofgren, and G. A. Gordon. 1979. Variations in climate since 1602 as reconstructed from tree rings. Quaternary Research **12**:18–46.

Fritts, H. C., and T. W. Swetnam. 1986. Dendroecology: a tool for evaluating variations in past and present forest environments. Laboratory of Tree-Ring Research, University of Arizona, Tucson, Arizona, USA.

Green, D. G. 1989. Simulated effects of fire, dispersal and spatial pattern on competition within forest mosaics. Vegetatio **82**:139–153.

Habeck, J. R. 1994. Impact of fire suppression on forest succession and fuel accumulations in long-fire-interval wilderness habitat types. USDA Forest Service General Technical Report INT-182. Rocky Mountain Research Station, Ogden, Utah, USA.

Hann, W. J., J. L. Jones, M. G. Karl, P. F. Hessburg, R. E. Keane, D. G. Long, J. P. Menakis, C. H. McNicoll, S. G. Leonard, R. A. Gravenmeir, and B. G. Smith. 1997. Landscape dynamics of the Basin. Pages 337–1056 *in* T. M. Quigley and S. J. Arbelheide, editors. An assessment of ecosystem components in the Interior Columbia Basin including portions of the Klamath and Great Basins. Volume 2. USDA Forest Service General Technical Report PNW-GTR-405. Pacific Northwest Research Station, Portland, Oregon, USA.

Hann, W. J., J. L. Jones, R. E. Keane, P. F. Hessburg, and R. A. Gravenmier. 1998. Landscape dynamics. Journal of Forestry **96**:10–15.

Hargrove, W. W., and R. J. Luxmoore. 1998. Ecoregions in the USA. Online: ⟨http://www.esd.ornl.gov/~hnw/esri98/⟩.

Haynes, R. W., R. T. Graham, and T. M. Quigley, technical editors. 1996. A framework for ecosystem management in the Interior Columbia Basin including portions of the Klamath and Great Basins. USDA Forest Service General Technical Report PNW-GTR-374. Pacific Northwest Research Station, Portland, Oregon, USA.

Hessburg, P. F., B. G. Smith, and S. G. Kreiter. 1999. Historical and current forest and range landscapes in the Interior Columbia River basin including portions of the Klamath and Great Basins. Part II: Linking vegetation patterns and landscape vulnerability to potential insect and pathogen disturbances. USDA Forest Service General Technical Report PNW-GTR-XXX. Pacific Northwest Research Station, Portland, Oregon, USA.

Heyerdahl, E. K., D. Berry, and J. K. Agee. 1995. Fire history database of the western United States. EPA Office of Research and Development: Report EPA/600/R-96/081. Washington, DC, USA.

Hobbs, R. J. 1998. Managing ecological systems and processes. Pages 459–485 *in* D. L. Peterson and V. T. Parker, editors. Ecological scale: theory and applications. Columbia University Press, New York, New York, USA.

Holdridge, L. R. 1947. Determination of world plant formations from simple climatic data. Science **105**:367–368.

Isaaks, E. H., and R. M. Srivastava. 1989. An introduction to applied geostatistics. Oxford University Press, New York, New York, USA.

Johnson, E. A., and S. L. Gutsell. 1994. Fire frequency models, methods, and interpretations. Advances in Ecological Research **25**:239–287.

Johnson, K. N., F. Swanson, M. Herring, and S. Greene, editors. 1999. Bioregional assessments: science at the crossroads of management and policy. Island Press, Washington, DC, USA.

Keane, R. E., S. F. Arno, and J. K. Brown. 1989. FIRESUM: an ecological process model for fire succession in western conifer forests. USDA Forest Service General Technical Report INT-266. Rocky Mountain Research Station, Ogden, Utah, USA.

Keane, R. E., and D. G. Long. 1998. A comparison of coarse-scale fire-effects simulation strategies. Northwest Science (special issue) **72**:76–90.

Keane, R. E., D. G. Long, J. P. Menakis, W. J. Hann, and C. D. Bevins. 1996*a*. Simulating coarse-scale vegetation dynamics using the Columbia River Basin Succession Model CRBSUM. USDA Forest Service General Technical Report INT-GTR-340. Rocky Mountain Research Station, Fort Collins, CO.

Keane, R. E., C. H. McNicoll, K. M. Schmidt, and J. L. Garner. 1996*b*. Spatially explicit ecological inventories for ecosystem management planning using gradient modeling and remote sensing. Proceedings of the Sixth USDA Forest Service Remote Sensing Applications Conference, Denver, Colorado, USA.

Keane, R. E., E. D. Reinhardt, and J. K. Brown. 1994. FOFEM: a first order fire effects model for predicting the immediate consequences of wildland fire in the United States. Pages 628–631 *in* Proceedings of the 12th Conference on Fire and Forest Meteorology, Society of American Foresters Publication 94–02. Jekyll Island, Georgia, USA.

Keane, R. E., K. Ryan, and S. W. Running. 1996*c*. FIRE-BGC—a mechanistic ecological process model for simulating fire succession on coniferous forest landscapes of the Northern Rocky Mountains. USDA Forest Service Research Paper INT-484. Rocky Mountain Research Station, Fort Collins, Colorado, USA.

Kercher, J. R., and M. C. Axelrod. 1984. A process model of fire ecology and succession in a mixed conifer forest. Ecology **65**:1725–1742.

Kessell, S. R. 1976. Gradient modeling: a new approach to fire modeling and wilderness resource management. Journal of Environmental Management **1**:39–48.

King, A. W., A. R. Johnson, and R. V. O'Neill. 1991. Transmutation and functional representation of heterogeneous landscapes. Landscape Ecology **5**:239–253.

Küchler, A. W. 1964. Potential natural vegetation of the coterminous United States. Special Publication 36 (with separate map at 1:3,168,000). American Geographical Society. New York, New York, USA.

Landres, P., P. Morgan, and F. J. Swanson. 1999. Overview of the use of natural variability concepts in managing ecological systems. Ecological Applications **9**:1179–1188.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm. Ecology **74**:1659–1673.

Lenihan, J., C. Daly, D. Bachelet, and R. P. Neilson. 1998. Simulating broad-scale fire severity in a dynamic global vegetation model. Northwest Science (special issue) **72**: 91–103.

Lertzman, K., J. Fall, and B. Dorner. 1998. Three kinds of heterogeneity in fire regimes: at the crossroads of fire history and landscape ecology. Northwest Science (special issue) **72**:4–23.

Madany, M. H., T. W. Swetnam, and N. E. West. 1982. Comparison of two approaches for determining fire dates from tree scars. Forest Science **28**:856–861.

Mathsoft Inc. 1994. Splus version 3. 3 for Windows and version 3.4 for UNIX. Seattle, Washington, USA.

McKenzie, D. 1998. Fire, vegetation, and scale: toward optimal models for the Pacific Northwest. Northwest Science (special issue) **72**:49–65.

McKenzie, D., D. L. Peterson, and E. Alvarado. 1996*a*. Predicting the effect of fire on large-scale vegetation patterns in North America. USDA Forest Service Research Paper PNW-489. Pacific Northwest Research Station, Portland, Oregon, USA.

McKenzie, D., D. L. Peterson, and E. Alvarado. 1996*b*. Extrapolation problems in modeling fire effects at large spatial scales: a review. International Journal of Wildland Fire **6**: 165–176.

Miller, C., and D. L. Urban. 1999. Interactions between forest heterogeneity and surface fire regimes in the southern Sierra Nevada. Canadian Journal of Forest Research **29**:202–212.

Morgan, P., G. H. Aplet, J. B. Haufler, H. C. Humphries, M. M. Moore, and W. D. Wilson. 1994. Historical range of variability: a useful tool for evaluating ecosystem change. Journal of Sustainable Forestry **2**:87–111.

Morgan, P., S. C. Bunting, A. E. Black, T. Merrill, and S. Barrett. 1996. Fire regimes in the interior Columbia River basin: past and present. Report on file at USDA Forest Service Intermountain Fire Sciences Laboratory, Rocky Mountain Research Station, Missoula, Montana, USA.

Morrison, P., and F. J. Swanson. 1990. Fire history and pattern in a Cascade Range landscape. USDA Forest Service General Technical Report PNW-254. Pacific Northwest Research Station, Portland, Oregon, USA.

Neilson, R. P. 1995. A model for predicting continental-scale vegetation distribution and water balance. Ecological Applications **5**:362–385.

Neter, J., W. Wasserman, and M. H. Kutner. 1990. Applied linear statistical models. Irwin Press, Homewood, Illinois, USA.

O'Neill, R. V. 1979. Natural variability as a source of error in model predictions. Pages 23–32 *in* G. S. Innis and R. V. O'Neill, editors. Systems analysis of ecosystems. International Cooperative Publishing House, Fairland, Maryland, USA.

O'Neill, R. V. 1998. Homage to St. Michael; or, why are there so many books on scale. Pages 3–17 *in* D. L. Peterson and V. T. Parker, editors. Ecological scale: theory and applications. Columbia University Press, New York, New York, USA.

Pahl-Wostl, C. 1995. The dynamic nature of ecosystems: chaos and order entwined. Wiley and Sons, New York, New York, USA.

Peterson, D. L., and K. C. Ryan. 1986. Modeling post-fire conifer mortality for long range planning. Environmental Management **10**:797–808.

Pickett, S. T. A., and J. N. Thompson. 1978. Patch dynamics and the design of nature reserves. Biological Conservation **13**:27–37.

Quattrochi, D. A., and M. F. Goodchild. 1997. Scale in remote sensing and GIS. CRC Press, Boca Raton, Florida, USA.

Quigley, T. M., R. W. Haynes, and R. T. Graham. 1996. Integrated scientific assessment for ecosystem management in the Interior Columbia Basin. USDA Forest Service General Technical Report PNW-GTR-382. Pacific Northwest Research Station, Portland, Oregon, USA.

Rastetter, E. B., A. W. King, B. J. Cosby, G. M. Hornberger, R. V. O'Neill, and J. E. Hobbie. 1992. Aggregating fine-scale ecological knowledge to model coarser scale attributes of ecosystems. Ecological Applications **2**:55–70.

Roberts, D. 1996. Landscape vegetation modelling with vital attributes and fuzzy systems theory. Ecological Modelling **90**:175–184.

Rossi, R. E., P. W. Borth, and J. J. Tollefson. 1993. Stochastic simulation for characterizing ecological spatial patterns and appraising risk. Ecological Applications **3**:719–735.

Rothermel, R. C. 1972. A mathematical model for predicting fire spread in wildland fuels. USDA Forest Service Research Paper INT-115. Intermountain Research Station, Ogden, Utah, USA.

Schmoldt, D. L., D. L. Peterson, R. E. Keane, J. M. Lenihan, D. McKenzie, D. R. Weise, and D. V. Sandberg. 1999. Assessing the effects of fire disturbance on ecosystems: a scientific agenda for research and management. USDA Forest Service General Technical Report PNW-GTR-455. Pacific Northwest Research Station, Portland, Oregon, USA.

Schmoldt, D. L., and H. M. Rauscher. 1995. Building knowledge-based systems for natural resource management. Chapman and Hall, New York, New York, USA.

Sierra Nevada Ecosystem Project (SNEP). 1996. Status of the Sierra Nevada. Final report to Congress. Centers for Water and Wildland Resources. University of California, Davis, California, USA.

Swetnam, T. W., C. D. Allen, and J. L. Betancourt. 1999. Applied historical ecology: using the past to manage for the future. Ecological Applications **9**:1189–1206.

Turner, M. G., R. H. Gardner, V. H. Dale, and R. V. O'Neill. 1989. Predicting the spread of disturbance across heterogeneous landscapes. Oikos **55**:121–129.

Turner, M. G., and W. H. Romme. 1994. Landscape dynamics in crown fire ecosystems. Landscape Ecology **9**:59–77.

van Wagner, C. E. 1977. Conditions for the start and spread of crown fire. Canadian Journal of Forest Research **7**:23–34.

van Wagner, C. E. 1993. Prediction of crown fire behavior in two stands of jack pine. Canadian Journal of Forest Research **23**:442–449.

Venables, W. N., and B. D. Ripley. 1994. Modern applied statistics with Splus. Springer-Verlag, New York, New York, USA.

Woodward, F. I. 1987. Climate and plant distribution. Cambridge University Press, London, UK.

Woodward, F. I., and I. F. McKee. 1991. Vegetation and climate. Environmental International **17**:535–546.